

Path-analysis of metric-first & entropy-first approaches

P. Fraundorf

*Physics & Astronomy/Center for NanoScience, U. Missouri-StL (63121) and
Physics, Washington University (63110), St. Louis, MO, USA**

(Dated: February 23, 2012)

Established idea-sets may not update seamlessly. The tension between new and old views of nature is documented in Galileo's dialogs and now present in many fields. One evolutionary response may be to consider the simplicity of paths from *various* starting points to one goal. We illustrate with a look at two simplifications: The move from **Lorentz-transform** to **metric-equation** descriptions of space-time, and the move from **classical** to **statistical thermodynamics** with help from Boltzmann's choice-multiplicity & Shannon's uncertainty. Connections of the latter to **correlation measures** behind available work, model selection, and layered complexity are also explored. New strategies are exemplified with Appendices on: anyspeed vector-velocity addition, the energy-momentum half-plane lost to finite lightspeed, the modern distinction between proper and geometric accelerations, single map-frame views of anyspeed acceleration, quantifying risk with a handful of coins, available work in bits, quantitative model-selection, and the evolution of analog/digital complexity.

PACS numbers: 05.70.Ce, 02.50.Wp, 75.10.Hk, 01.55.+b

Contents

I. Introduction	1
II. Principles	1
III. Metric-based motion	2
IV. Multiplicity-based thermodynamics	3
V. Collateral connections	3
A. surprisals	3
B. average surprisals	4
C. net surprisals	5
VI. Discussion	6
VII. Conclusions	6
Acknowledgments	6
A. Proper-velocity	6
1. vector addition	6
2. energy vs. momentum	7
B. Proper-acceleration	7
1. proper or geometric?	7
2. accelerated roundtrips	8
C. Choice multiplicities	8
1. quantifying risk	9
D. Matchup multiplicities	9
1. work's availability	9
2. model selection math	10
3. layered correlations	11
References	12

I. INTRODUCTION

Evolution of well-worn approaches naturally encounters resistance from experts in the old¹. For instance Martin Gardner in his book on parity inversion² cites Hermann Kolbe's negative reaction to the prediction of carbon's tetrahedral nature by Jacobus van't Hoff (Chemistry's first Nobel Laureate). The hullabaloo³ about the Nowak et al. paper⁴ on models for evolving insect social behavior is a more recent research example, while participants in the content-modernization branch of physics education research (PER) have engaging tales on the education side⁵. For text publishers, however, even funerals may not mark progress since choosers of a course text might understandably like to teach that course the way they learned it, whether they own the strategy or not.

One way to objectively assess new approaches is perhaps to examine the algorithmically-shortest path to quantitative insight from each given starting point. For experts in the old, traditional approaches may be algorithmically-shortest even if they are not shortest for newcomers to a given subject. Differing perceptions, in this context, might thus be put onto a rational footing. In this context here, we examine textbook trends toward "metric-first" approaches to relativistic motion, and "entropy-first" approaches to statistical inference about physical systems, in hopes of helping individual teachers chart their own path through the evolving terrain.

II. PRINCIPLES

From a given starting point, the strategy for putting together a concept map may be to minimize the number of: (i) assumptions and (ii) new concepts needed to make a given set of quantitative predictions possible. Drawings

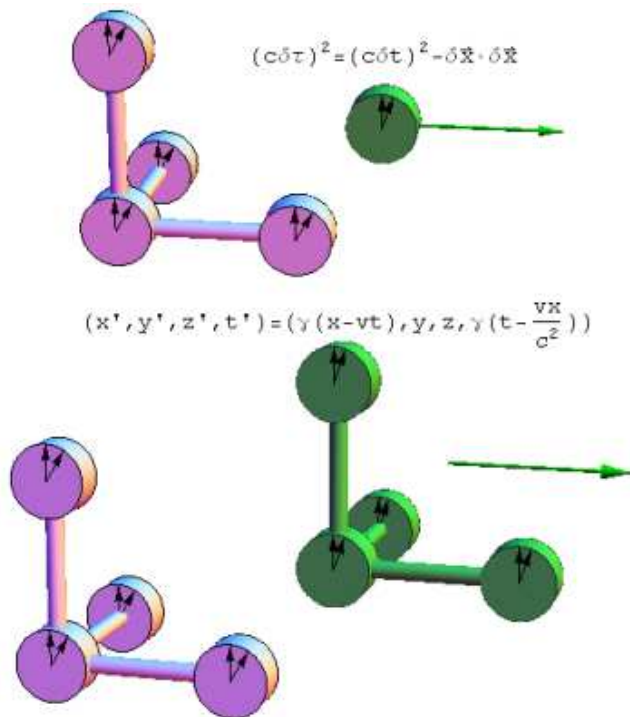


FIG. 1: Proper-kinematics from the metric-equation first.

of such paths from various starting points, in this context, might inspire one to evolve one's own starting point in teaching a given class over time.

Note that we are weighing self-consistent approaches for their compactness, portability, and appropriateness in much the same way that different variable-changes in calculus, and coordinate-system choices in analytic geometry, buy more advantage for some tasks and less advantage for others. In that sense, we seek to apply the science of *Bayesian model-selection* to the evolution of what we teach.

Below we illustrate this with a few examples, based on content changes already underway in the evolving physics curriculum. Similar charts for your own approach to a given class, as it relates to the textbook in hand as well as the larger picture, may be worth putting together for sharing with your students and perhaps in electronic collaboration spaces with the larger physics teaching community as well.

III. METRIC-BASED MOTION

The traditional path via *Lorentz transforms*, by using two separate coordinate-frames with their own yardsticks and synchronized clocks, likely provides the most direct route to **length contraction**. On the other hand shorter paths to **time dilation**⁶, **accelerated motion**⁷, and **gravitation**^{8,9} follow by invoking the flat-space *metric equation* to a single map-frame of yardsticks & synchro-

nized clocks, as simply an equation for time-elapsd on the clocks of a traveler.

The shift may seem uninteresting if one has mastered space-time through Lorentz transforms, and sees relativity as an extension of Newtonian physics to extreme situations. On the other hand if one has been exposed mainly to single map-frame calculations and/or sees relativity as an everyday cause of magnetism & gravity, then the two-frame approach serve up un-needed complication.

The traditional approach for instance: (i) emphasizes symmetry between frames even when the home-frame e.g. of a traveling clock or yardstick is quite special, (ii) raises the dissonant spectre^{10–13} of relativistic mass, (iii) avoids use of proper-acceleration^{14,15} as an integrative complement to geometric-accelerations (affine-connection effects) at low and high speeds, and (iv) misses out on insights that proper-velocity^{16,17} offers e.g. into relativistic velocity-addition (Appendix A 1) and the lightspeed limit (Appendix A 2). The single map-frame approach avoids these problems, with the result that elements of it are finding their way into texts on all levels.

In context of this evolving strategy we recommend drawing concept-interconnection maps (like those in the sections below) showing the steps from what you, and what separately your students, understand to what you'd like them to master in a given course. Intro-physics and relativity texts over the past couple of decades will give you an idea about how the Lorentz-transform approach is already evolving to consider privileged frames e.g. like that of the traveling clock in time-dilation analysis and the traveling yardstick in length-contraction analysis. Because it doesn't yet appear in texts, however, let's show briefly how a metric-first approach can build significant insight from what intro-physics students already know.

Begin with Minkowski's version of Pythagoras theorem i.e. the flat-space metric equation shown in the top of Figure 1. This introduces time-elapsd on the traveler's clock i.e. proper-time τ , and hence two new rate-of-travel parameters: namely the 3-vector proper-velocity $\vec{w} \equiv d\vec{x}/d\tau$ and Lorentz-factor $\gamma \equiv dt/d\tau$ in terms of already-familiar coordinate-velocity $\vec{v} \equiv d\vec{x}/dt$.

For anyspeed work, proper-velocity instead of coordinate-velocity: (i) equals momentum per unit mass, (ii) adds vectorially (with an out-of-frame rescale) as shown in Appendix A 1, and (iii) with no upper-limit most elegantly (at 1 [ly/ty]) parameterizes the transition from sub to hyper relativistic. Moreover this metric-first analysis can be extended to treat constant proper-acceleration with equations that in the unidirectional case are particularly simple, and connected to the low-speed equations for constant coordinate-acceleration.

The big caveat here is that simultaneity is defined entirely by the single reference map-frame. Robust quantitative insight into the meaning of that, as well as of length-contraction, likely will have to await Lorentz transforms.

As teachers we should probably choose a path through

space-time for students which draws strength from our prior training and acquired insight into both paths, as well as the path's connection to the past and future of students in each given course. For instance, with introductory physics students it's quite easy to tell students (even if the book doesn't) that time passes differently on different clocks so that, unless otherwise stated, time will be measured on a set of synchronized "map-clocks" affixed to the yardsticks used to measure position. Even better if we can also give them an updated view of forces in non-rain frames (Appendix B 1), and of constant acceleration at any speed (Appendix B 2).

IV. MULTIPLICITY-BASED THERMODYNAMICS

The question here is: Do I start by introducing temperature in historical units and the zeroth law while saving entropy to the end, or do I start with choice-multiplicity and entropy so that the assumptions behind the ideal gas law, equipartition, and mass action are explicit from day one? Senior undergraduate texts almost all now do the latter, while only a small number of introductory texts have made the switch so far.

The simplest axiomatic path to: (a) the ideal gas equation, (b) equipartition, and (c) the law of mass action is likely Boltzmann's choice-multiplicity

$$W = \prod_{i=1}^N \left(\frac{1}{p_i} \right)^{p_i}, \quad (1)$$

from which entropy $S = k \ln[W]$ and its derivatives may be defined. This choice-multiplicity, of course, is just the dimensionless count which underlies the familiar use of both information units and Joules per degree Kelvin[12].

Historical approaches introduce these consequences as empirical and/or informally-useful relationships, without clear definition of their underlying mechanism and assumptions and typically with discussions of entropy/multiplicity (the horse) following these consequences (the cart). Such approaches do not provide insight into: (i) the quantitative limitations of these concepts, or (ii) strategies for moving beyond those limitations e.g. to systems in which subsystem correlations cannot be ignored.

Another reason to introduce multiplicity first is that the laws of thermodynamics (short of two physical postulates) follow therefrom as well. The zeroth law follows from the fact that the largest number of states is available when the uncertainty slopes of two subsystems (reciprocal temperatures for the energy observable) equilibrate as a conserved quantity is shared.

The first law, oft described as a statement of energy conservation, in fact arises from maximum entropy inference as a relation between ordered and disordered changes in any observable, whether they are conserved on transfer between subsystems or not. Likewise for the

second law, whose physics actually comes not from statistical inference but from the assumption that mutual information available on the state of an isolated system will not increase over time.

Finally the very definition of reciprocal temperature as an uncertainty slope will convince many that the change in state-uncertainty about any finite system, per unit change in energy, is likely to be finite. Hence reciprocal-temperature's infinity (the absolute-zero of temperature) is likely inaccessible. This natural definition of temperature has the added advantage that it prohibits one from approaching absolute-zero from negative or positive directions, and shows that the negative absolute-zero approachable e.g. by spin systems with a population inversion is as far away from positive absolute-zero as you can get.

Examples of the power in this recasting of familiar rules include many senior undergraduate thermal physics texts, like those by Kittel & Kroemer¹⁸, Dan Schroeder¹⁹, and Claude Garrod²⁰ (who refers to reciprocal temperature²¹ as *coldness*), Tom Moore & Dan Schroeder's AJP paper²², Tom Moore's introductory physics Unit T⁶, etc.

V. COLLATERAL CONNECTIONS

We've now covered two paradigm-shifts that have a well-defined place in the physics curriculum. The approach taken with respect to them in a given class should inform itself to both teacher & student backgrounds, as well as to *course objectives*. The second paradigm-shift makes contact with other developments of interest to physics students as well.

To explore this we step back from uncertainties to probability measures, and then forward from uncertainties to correlation measures, to show how the second simplification also allows physics to make contact with a number of other lively disciplines. Because of the physics in between, out-of-discipline students may never hear about these connections if they aren't at least mentioned in one of their physics classes.

A. surprisals

Recall that information units can be introduced by the statement that # choices equals $2^{\# \text{bits}}$. Also very small probabilities p can be put into everyday terms as the **surprisal**²³ $s = n$ bits of tossing n coins all heads up since $p = 1/2^{\# \text{bits}}$, with the added advantage that surprisals add whenever their probabilities multiply (Appendix C 1). Evidence in bits²⁴ for a true-false proposition can similarly be written as $e[p] = s[1-p] - s[p]$, where surprisal is $s[p] = \ln_2[1/p]$.

All of these applications rely on the fact that probabilities between 0 and 1 can be written as multiplicities $w_p = 1/p$ between 1 and $+\infty$ or as surprisals between 0

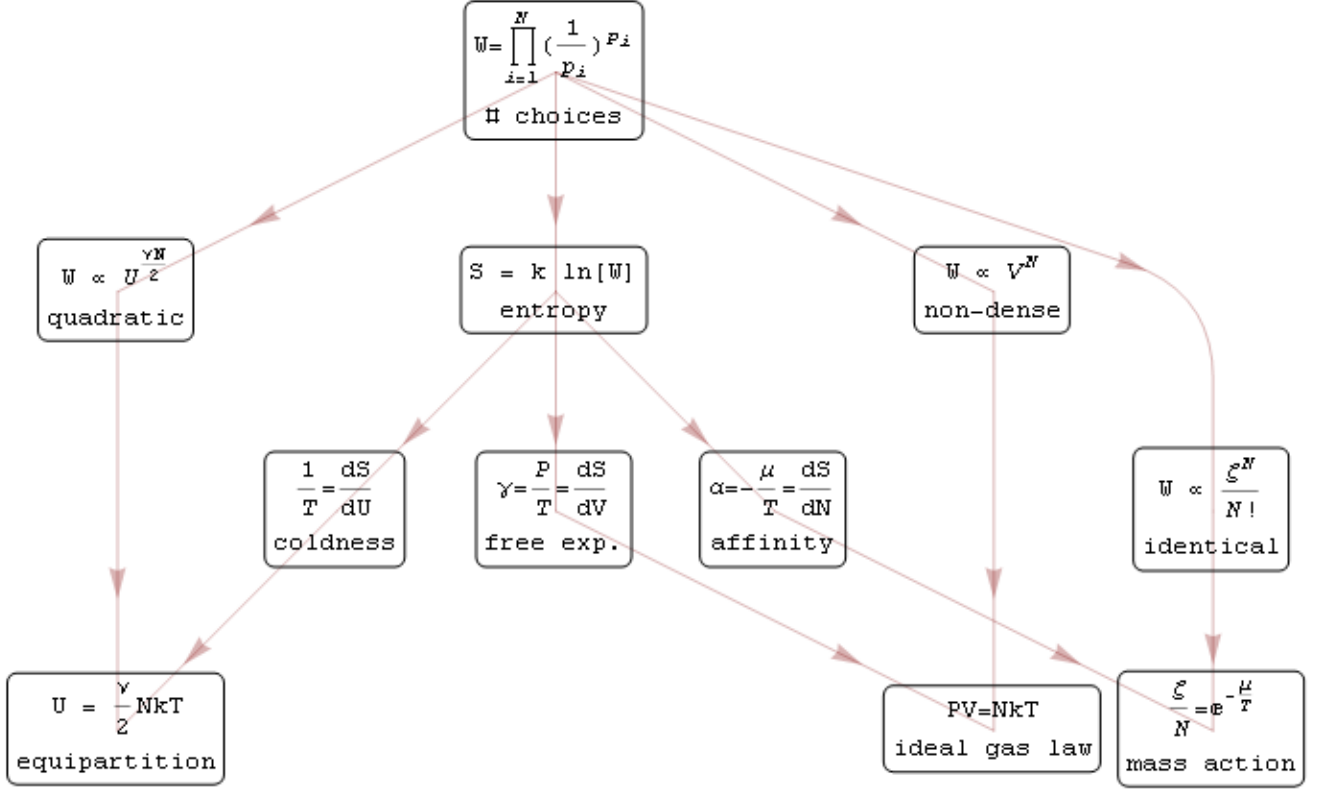


FIG. 2: Choice multiplicity \Rightarrow gas law, equipartition & mass action.

and $+\infty$ using information units determined by the constant k in the expression $s_p = k \ln[1/p]$. This surprisal \Leftrightarrow multiplicity \Leftrightarrow probability inter-conversion is summarized by:

$$0 \leq s_p \equiv k \ln[w_p] \equiv k \ln\left[\frac{1}{p}\right] \leq \infty \quad (2)$$

where of course the units are bits if $k = 1/\ln[2]$.

B. average surprisals

The treatments of the ideal gas law, equipartition, mass action, and the laws of thermodynamics in the previous section connect to this tradition by defining uncertainty or entropy S as an **average surprisal** e.g. in J/K between 0 and $+\infty$, Boltzmann's multiplicity W between 1 and $+\infty$ as $e^{S/k}$ where k is Boltzmann's constant, and $1/W$ as a reciprocal-multiplicity between 0 and 1. Their relevance to the thermal side of physics education has been discussed above.

More generally the interconversion for the average surprisal, uncertainty, or entropy associated with predicted probability-set q , as measured by operating probability-

set p , can be written:

$$0 \leq S_{p/p} \leq S_{q/p} \equiv k \ln[W_{q/p}] \equiv k \sum_{i=1}^N p_i \ln\left[\frac{1}{q_i}\right] \leq \infty. \quad (3)$$

Thus $S_{q/p}$ e.g. for an observation is in bits the **average-surprisal if the expected model q differs from the operating-model p** .

Although written for a discrete probability-set, the expression is naturally adapted to continuous as well as quantum mechanical probability-sets^{25,26}. In this context natural as distinct from historical units for temperature become energy per unit information, and for heat capacity become bits²⁷.

Note that the upper limit on $S_{p/p}$ is $\ln_2[N]$. Also the fact that $S_{q/p} \leq S_{p/p}$, i.e. that **measurements using the wrong model q are always likely to be more surprised by observational data than those using the operating-model p** , underlies maximum-likelihood curve-fitting and Bayesian model-selection as well as the positivity of the correlation and thermodynamic availability measures discussed below.

Thus in this two-distribution case, $1 \leq W_{q/p} \leq +\infty$ is an **effective choice-multiplicity** for expected probability set q in the face of operating-probability set p . In general $W_{p/p} \leq W_{q/p}$. For the uniform N -probability set $u_i = 1/N$ for i running from 1 to N , we can also say that

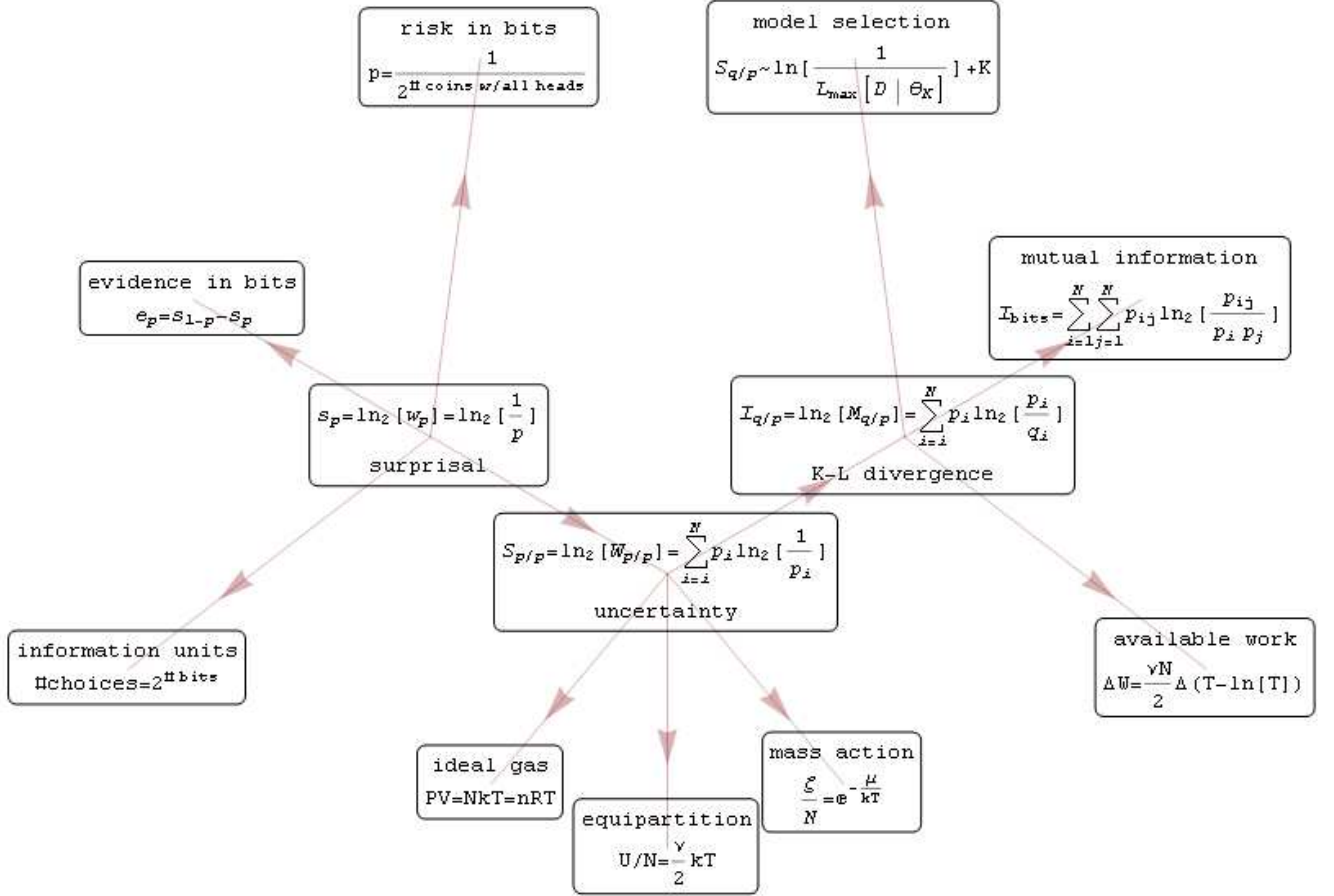


FIG. 3: Cross-disciplinary applications for log-probability measures in statistical inference.

$$W_{p/p} \leq W_{u/p} = N = W_{u/u} \leq W_{p/u}.$$

C. net surprisals

The tracking of subsystem correlations has taken a back seat in traditional thermodynamic use of log-probability measures. This is illustrated e.g. by the traditional treatment of subsystem entropies as additive, in effect promising that correlations between e.g. between gas atoms in two volumes separated by a barrier can be safely ignored. More generally, however, subsystem correlations (e.g. between a sent and a received message, or between traits of a parent and of a child) are of central importance. In fact the maximum entropy discussed above is nothing more than minimum KL-divergence with a uniform prior²⁸, so that physicists expert in its application to analog systems can play a pivotal role informing students who take physics courses about these connections across disciplines.

In particular the foregoing are backdrop to the paradigm-shift which broke out of physics into the wide world of statistical inference in the mid-20th century²⁹. We'll touch on only three of the many areas that it's

connecting together today, based on their relevance to cross-disciplinary interests of students in physics classes. The specific application areas are: (i) **thermodynamic availability** as in Appendix D 1, (ii) algorithmic **model selection** as in Appendix D 2, and (iii) the **evolution of complexity** as in Appendix D 3. The surprisal \Leftrightarrow multiplicity \Leftrightarrow probability interconversion for these correlation analyses may be written:

$$0 \leq I_{q/p} \equiv k \ln [M_{q/p}] \equiv k \sum_{i=1}^N p_i \ln \left[\frac{p_i}{q_i} \right] \leq \infty \quad (4)$$

Log-probability measures are useful for tracking subsystem-correlations in digital as well in analog complex systems. In particular tools based on Kullback-Leibler divergence $I_{q/p} \geq 0$ (the negative of Shannon-Jaynes entropy) and the matchup-multiplicity or choice-reduction-factor $M_{q/p}$ associated with reference probability-set q have proven useful: (i) to engineers for measuring available-work or *exergy* in thermodynamic systems³⁰, (ii) to communication scientists and geneticists for studies of: regulatory-protein binding-site structure³¹, relatedness³², network structure, & replication fidelity^{33,34}, and (iii) to behavioral ecologists want-

ing to select from a set of simple-models the one which is least surprised by experimental data^{35,36} from a complex-reality.

In context of this idea-set, the logical schematic in Figure 3 illustrates connections that often go unmentioned between what are now-classical application areas in their specialized fields. It thus suggests that physicists, particularly thanks to their long experience with log-probability measures in analog systems, can play a key role in the cross-disciplinary application of informatics to complex systems.

These multi-moment correlation-measures have 2nd law teeth making them relevant to quantum computing³⁷, and they enable one to distinguish pair from higher-order correlations making them relevant to the exploration of order-emergence in a wide range of biological systems^{38,39}. They may be especially useful in addressing challenges associated with the sustainability of multi-layer complex systems⁴⁰.

VI. DISCUSSION

Similar analyses might also help *each of us* decide when it is (and is not) appropriate to spend time in the educational arena e.g. on: (i) geometric-algebra approaches^{41–43} to complex numbers & cross-products, (ii) energy⁶ & least-action⁴⁴ based introductions to mechanics, (iii) vector potential introductions to magnetism⁴⁵, (iv) explore-all-paths introductions to quantum mechanics⁴⁶. The approach may even come in handy for mediating differences in research strategy as well, e.g. in deciding how much time to spend (in context of a particular problem) on: (a) CPT approaches to the application of non-Hermitian Hamiltonians⁴⁷, (b) molecule-code as distinct from kin-selection models of evolving eusocial or altruistic behavior⁴, etc.

VII. CONCLUSIONS

In short both quantitative and schematic considerations of the algorithmic path to key deliverables from **your & your audience's** conceptual starting point may help point you toward approaches that help your students become maximally-informed in minimum time. These may not lessen “the detailed work of content modernization”⁵, but they may help provide the process with useful direction taylorred to our individual points of reference. What would your concept maps look like in this context?

Acknowledgments

I would like to thank graduate students Bob Collins, Zak Jost, and Pat Sheehan for their participation in a

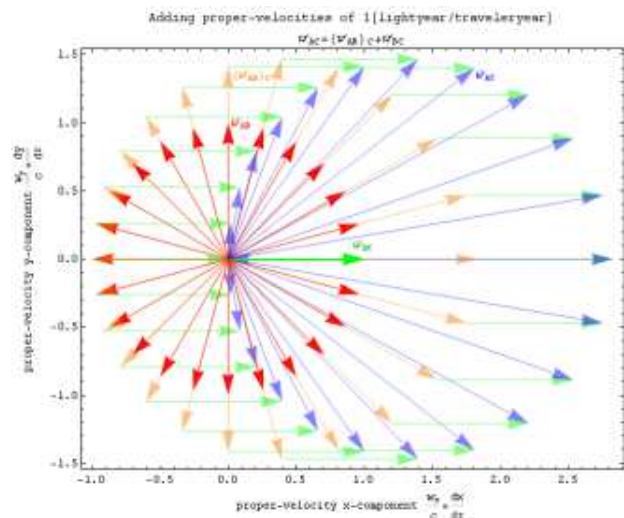


FIG. 4: Adding $w/c = 1$ proper-velocity vectors.

Bayesian informatics development course which helped to nail down some of the inter-connections discussed here.

Appendix A: Proper-velocity

Minkowski’s flatspace metric-equation naturally defines the relation between traveler-time elapsed ($\delta\tau$) and the distance/time between events defined with respect to the yardsticks ($\delta\vec{x}$) and synchronized-clocks (δt) of a single map-frame, thus defining Lorentz factor & proper-velocity as alternate ways to describe rate of travel.

Proper-velocity, referred to by Shurcliff as the “minimally-variant” parameter for describing position’s rate of change, can simplify our understanding of many relativistic processes. We choose two ways here that are relevant to an introductory physics course.

1. vector addition

A useful mnemonic for relative motion in the Newtonian world is:

$$\vec{v}_{AC} = \vec{v}_{AB} + \vec{v}_{BC} \quad (\text{A1})$$

where e.g. \vec{v}_{AB} is the vector velocity of object A with respect to object B . Note that in general $\vec{v}_{AB} = -\vec{v}_{BA}$, and in sums “a common middle letter cancels out”.

For a relation that works for *uni-directional* velocity-addition even at coordinate-speeds v near lightspeed c , one might use the similar relationship:

$$w_{AC} \equiv \gamma_{AC} v_{AC} = \gamma_{AB} \gamma_{BC} (v_{AB} + v_{BC}) \quad (\text{A2})$$

where the proper-velocity $\vec{w} \equiv d\vec{x}/d\tau = \vec{p}/m$ is map-distance \vec{x} traveled per unit time τ on traveler-clocks, coordinate-velocity $\vec{v} \equiv d\vec{x}/dt$ with $v \leq c$ is the usual

TABLE I: Acceleration types and various forces.
force name type | proper | geometric: non-free | non-local

normal	⊕		
string	⊕		
spring	⊕		
friction	⊕		
drag	⊕		
centripetal	⊕		
electromagnetic	⊕		
gravity		⊕	
reaction “gees”		⊕	
centrifugal		⊕	
Coriolis effect			⊕
tidal			⊕

equivalence principle which guarantees that Newton’s Laws can be helpful *locally* in accelerated frames and curved space time, provided that we invoke **inertial forces** to explain the geometric-accelerations which operate in those frames.

The mathematics of geometric accelerations comes from the fact that in *general relativity* an object’s coordinate acceleration (as distinct from only its proper-acceleration 4-vector A) is equal to:

$$\frac{dU^\lambda}{d\tau} = A^\lambda - \Gamma^\lambda_{\mu\nu} U^\mu U^\nu \quad (B1)$$

where geometric-accelerations are represented by the affine-connection term Γ on the right hand side. These may be the sum of as many as sixteen separate velocity and position dependent terms. Coordinate acceleration goes to zero whenever proper-acceleration is exactly canceled by that connection term, and thus when physical and inertial forces add to zero.

2. accelerated roundtrips

For unidirectional (1+1)D motion, the rapidity or hyperbolic velocity angle η simply connects the interchangeable velocity parameters Lorentz-factor $\gamma \equiv dt/d\tau$, proper-velocity $w \equiv dx/d\tau$ and coordinate-velocity $v \equiv dx/dt$ via:

$$\eta \equiv \sinh^{-1} \left[\frac{w}{c} \right] = \tanh^{-1} \left[\frac{v}{c} \right] = \pm \cosh^{-1} [\gamma] \quad (B2)$$

These parameters may then be used to express the proper-acceleration α experienced by an object traveling with respect to a map-frame of co-moving yardsticks and synchronized clocks in flat space time, in terms of its coordinate-acceleration a which cannot be held constant at high speed, as:

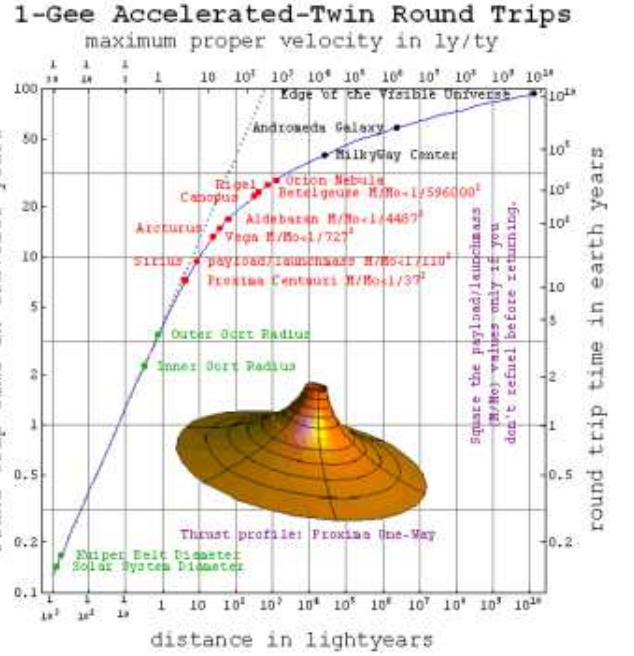


FIG. 6: 1-gee proper-acceleration roundtrips.

$$\alpha \equiv \frac{1}{\gamma} \frac{dw}{d\tau} = \gamma^3 a, \text{ where } a \equiv \frac{dv}{dt} \quad (B3)$$

This yields three integrals of constant proper-accelerated motion that reduce to the familiar equations of constant coordinate-acceleration at low speeds:

$$\alpha = \frac{\Delta w}{\Delta t} = c \frac{\Delta \eta}{\Delta \tau} = c^2 \frac{\Delta \gamma}{\Delta x} \quad v \ll c \quad \frac{\Delta v}{\Delta t} \quad v \ll c \quad \frac{1}{2} \frac{\Delta(v^2)}{\Delta x} \quad (B4)$$

These in turn allow one for example to write out analytical solutions (cf. Fig. 6) for round-trips involving constant 1 gee $\simeq 1.03[\text{ly}/\text{y}^2]$ accelerated/decelerated travel between stars.

Appendix C: Choice multiplicities

Senior physics courses have for already been re-arranged considering the fundamental role that multiplicity (and its logarithm, namely entropy) play in understanding and predicting behaviors. Although intro-physics courses are weaker in this context, books like Tom Moore’s “Six Ideas”⁶ have put choice-multiplicity where it belongs at the start of the thermo-chapters.

Hence the only section in this Appendix is one for students with virtually no math background. The hope is that teachers will individually explore ways to introduce the connection between bits and J/K, while at the same time nurturing an appetite for textbook revisions that better communicate the relation between thermal physics and information theory downstream.

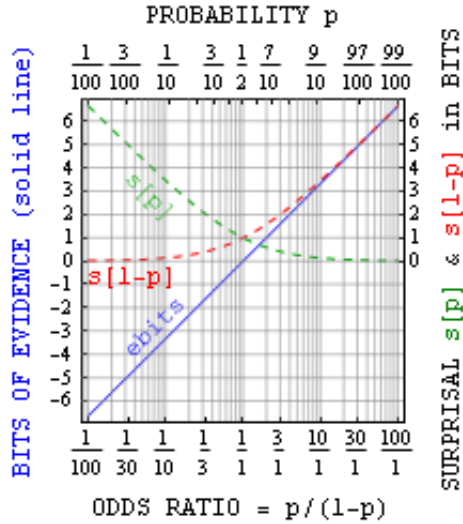


FIG. 7: Logarithmic measures of probability and odds.

1. quantifying risk

Surprisal in bits (defined by $\text{probability} = 1/2^{\text{#bits}}$) might be useful to citizens in assessing risk and/or standards of evidence (cf. Fig. 7), because of its simple, intuitive, and testable ability to connect even very small probabilities with one's experience at tossing coins. For example, the surprisal of dying from a smallpox vaccination (one in a million) is about 19.9 bits (like 20 heads in 20 tosses), while the surprisal of dying from smallpox once you have it (one in three) is only about 1.6 bits (i.e. more likely than 2 heads in 2 tosses).

Thus surprisal: (i) has meaning which is easy to remind yourself of with a few coins in your pocket, (ii) reduces huge numbers to much more intuitive size, and (iii) allows one to combine risks "from independent events" with addition/subtraction rather than multiplication/division.

For instance (from the numbers suggested above) your chance of dying is decreased by getting the vaccination, as long as the surprisal of getting smallpox without the vaccination is less than $20 - 2 \simeq 18$ bits. That means that vaccination is your best bet (absent other information) if your chances of being exposed to smallpox are greater than those of getting 18 heads in 18 tosses (1 out of $2^{18} \simeq 333,333$).

Given the large difference between something with 2 bits of surprisal and something with 18, communications bandwidth might be better spent by newsmedia providing us with numbers on observed surprisal, rather than by reporting only that "there's a chance" of something bad (or good) happening. Saying the latter treats your audience as consumers of spin rather than information.

Likewise, use of surprisals in communicating and monitoring risks to medical patients could make patient decisions about actions with a small chance of dire outcomes as informed as possible. This could reduce the

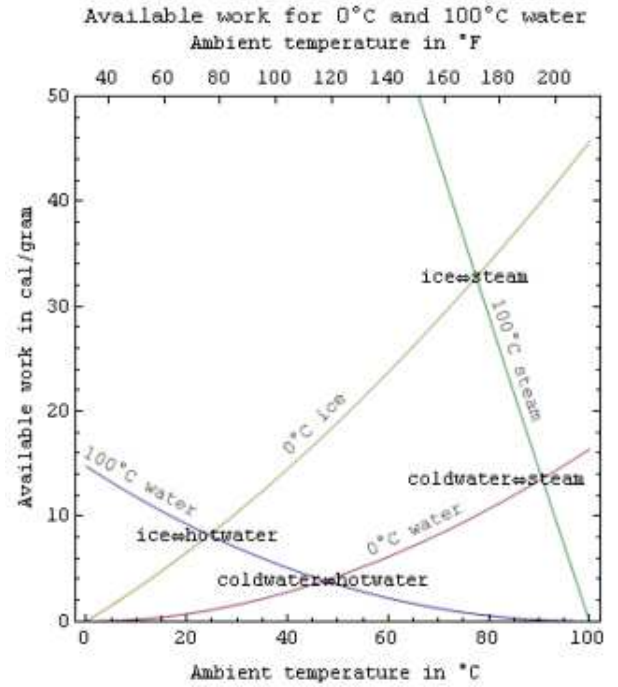


FIG. 8: Available work/gram vs. ambient temperature T_0 for 0° and 100° water in various states.

costs of medical malpractice in the long run by empowering patients with tools to make informed and responsible choices, making the need for legal redress less frequent.

Thus the media for risk-assessing public could play a key role in reducing the costs of defensive medicine. Some might even enjoy surprisal data on the small probabilities associated with some gambling opportunities. After all, there really is more to the lottery than simply knowing "the size of the pot".

Appendix D: Matchup multiplicities

Of the integrative concepts discussed in this paper, the least familiar to physicists (judging from textbooks, at least) may be those associated with the logarithmic correlation-measure often referred to as KL-divergence and its multiplicity: in effect a kind of normalized **choice-reduction factor** which is never less than 1. Hence in this Appendix we discuss matchup-multiplicity connections to: (i) available work, (ii) model-selection math important *at least* for physicists involved in cross-disciplinary work, and (iii) the evolution of complex analog as well as digital systems.

1. work's availability

Best-guess states (e.g. for atoms in a gas) are inferred by maximizing the average-surprisal S (entropy) for a

given set of control parameters (like pressure P or volume V). This constrained entropy maximization, both classically and quantum mechanically, minimizes Gibbs availability in entropy units $A = -k \ln[Z]$ where Z is a constrained multiplicity or partition function.

When absolute temperature T is fixed, free-energy (T times A) is also minimized. Thus if T , V and number of molecules N are constant, the Helmholtz free energy $F = U - TS$ (where U is energy) is minimized as a system “equilibrates”. If T and P are held constant (say during processes in your body), the Gibbs free energy $G = U + PV - TS$ is minimized instead. The change in free energy under these conditions is a measure of available work that might be done in the process. Thus available work for an ideal gas at constant temperature T_o and pressure P_o is $W = \Delta G = NkT_o\Theta[V/V_o]$ where $V_o = NkT_o/P_o$ and by Gibbs inequality $\Theta[x] \equiv x - 1 - \ln[x] \geq 0$.

More generally, the work available relative to some ambient is obtained by multiplying ambient temperature T_o by KL-divergence or net-surprisal $\Delta I \geq 0$, defined as the average value of $k \ln[p/p_o]$ where p_o is the probability of a given state under ambient conditions. For instance, the work available in equilibrating a monatomic ideal gas to ambient values of V_o and T_o is thus $W = T_o\Delta I$, where KL-divergence $\Delta I = Nk(\Theta[V/V_o] + (3/2)\Theta[T/T_o])$. The resulting contours of constant KL-divergence put limits on the conversion of hot to cold as in flame-powered air-conditioning, or in the unpowered device to convert boiling-water to ice-water (Fig. 8). Thus KL-divergence (also known to engineers as available-work or *exergy* in units of kT_o) measures thermodynamic availability in bits.

2. model selection math

In spite of: (i) the central role of model-selection in all observational science, and (ii) the math background required to do physics, the mathematics of model-selection is sometimes hardly an afterthought in physics tests and assessments. Introductory physics texts sometimes even define models as simplified *versions* of physical systems⁴⁸, instead of as idea-based representations that (like the molecule-based code-strings of one-celled organisms) help us correlate our behaviors with the world around.

Hence the background of physicists, e.g. on dissertation committees of biophysics students whose project-literature requires a mathematical approach to model-selection, may lead them to think that parameter-estimation and model-selection are one and the same. In this section we follow astronomer Phil Gregory’s discussion of Occam factors to show how the two relate, and point to an interesting strategy for getting quick answers that so far seems to be better known in ecology than in physics.

We begin with Gregory’s Occam factor²⁸, defined as $\Omega_M \equiv p[D|A, M]/p[D|\hat{A}, M]$ i.e. as the factor by which the likelihood of a set of data D is increased by consid-

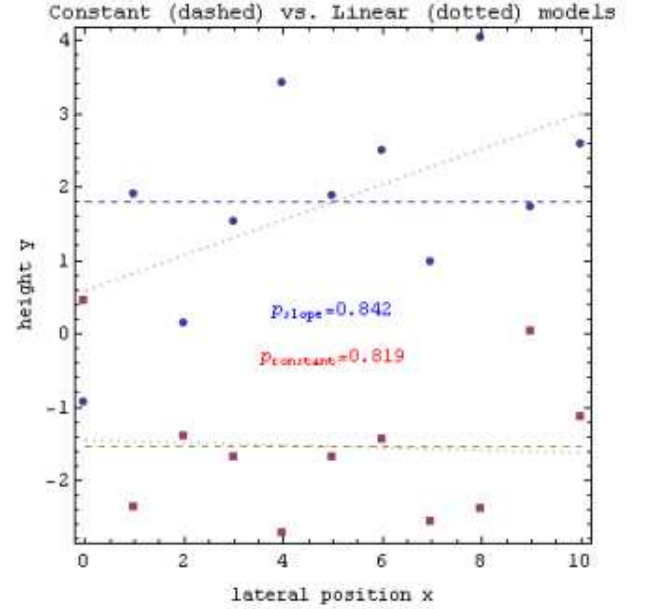


FIG. 9: Probabilities of constant vs. linear fits using AIC.

ering only the most likely set $\{\hat{A}_k\}$ of $k = 1$ to K fit-parameter values $\{A_k\}$ associated with a given model M . One might then use this to predict the surprisal of data generated by reality’s unknown operating probability q , namely $S_{p/q} \equiv \langle \ln \left[\frac{1}{p} \right] \rangle_q$ given that model:

$$S_{p/q} \simeq \ln \left[\frac{1}{p[D|A, M]} \right] = \ln \left[\frac{1}{p[D|\hat{A}, M]} \right] + \ln \left[\frac{1}{\Omega_M} \right] \quad (D1)$$

Although we can’t calculate KL-divergence $I_{p/q}$ directly because we don’t know $S_{q/q}$, the argument is that the best model M available (in the absence of inside information about reality q) is the one that is least surprised by available data D i.e. for which $S_{p/q}$ is least.

Note that the first term in the minimized quantity on the right-hand-side of Equation D1 is just the negative log-likelihood of the fit i.e. the quantity traditionally minimized when choosing the optimum set of parameters for a given model. For example, that first term when least-squares fitting N data points with normally-distributed errors having a mean μ and standard deviation σ is a non-varying constant plus a second constant times the average square of data deviations from the mean i.e. the model’s variance or mean-square-error.

When choosing one model over another, however, the second “Occam-factor” term must also be considered in the analysis. Gregory provides a lovely general expression for Occam factors in the linear least-squares case, in terms of prior probabilities for the fit parameters and the parameter-function covariance matrix. On the other hand statisticians have set-up simpler rules-of-thumb for this Occam factor, also grounded in Bayesian inference and the connection described here to KL-divergence.

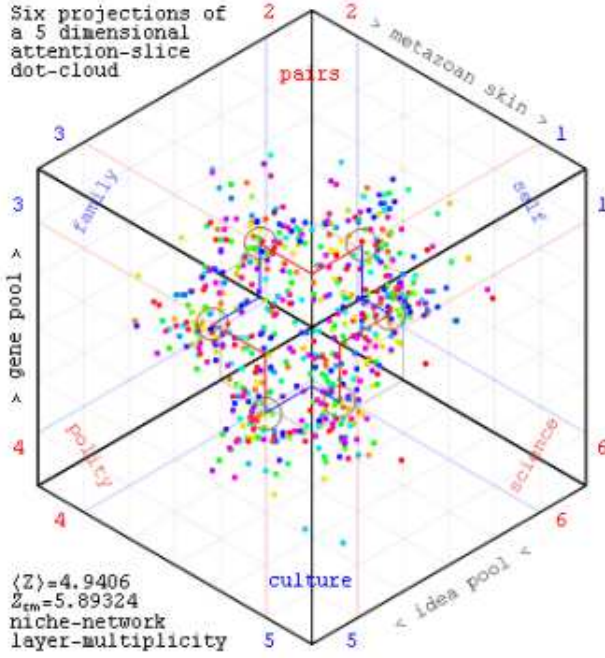


FIG. 10: Each community member shows up as a dot, in each of the six triangular projections.

Akaike information criterion (AIC) may be the best-documented^{35,36} of these rules of thumb. Although it traditionally uses an information-units constant of 2 instead of 1 for nats and $1/\ln[2]$ for bits, AIC given enough data basically replaces the Occam-factor term on the right-hand-side of Equation D1 with model-penalty of K nats, where K is the number of fit-parameters in the model. Use of AIC to decide if two sets of data points are better modeled with a sloped-line ($K=2$), or with a constant ($K=1$), is illustrated in Figure 9.

3. layered correlations

Does it seem strange that the second law of thermodynamics is expressed in terms of some observer's knowledge about a system's state? This strangeness is exacerbated by the fact that many of the analog systems that thermodynamics was traditionally applied to (like macroscopic amounts of gas and condensed matter) allow one to treat entropy as an extensive state-variable, by neglecting correlations between subsystems.

More generally of course $S_A + S_B - S_{AB} \geq 0$ with the difference being none other than the mutual information, i.e. the KL-divergence $I_{q/p}$ between the operating set of probabilities p and the product of marginal probabilities q taken for the two systems separately⁴⁹. With nano-sized systems, as well as with systems involving digital codes (built from either ideas or molecules), these corre-

lations can no longer be ignored.

Of course these correlations are also not being ignored by the second law, which in effect treats knowledge (residing outside a system) concerning a system's detailed state as a subsystem correlation. This correlation can be lost by accidental alterations to either the system or to the knowledge-repository, so that thermodynamic information (like any correlation between sub-systems) is a *delocalized* physical quantity.

Such sub-system correlations (including exergy over kT_o as discussed in Appendix D1) naturally evolve over time when a stream of ordered-energy (e.g. in the form of 2 eV solar photons) is available. In some cases, layered-correlations, that look in and out from a hierarchy of sub-system boundaries, manage to evolve as well⁵⁰.

Physical layer-boundaries emerge with the institutionalization of broken symmetries, as in the case of simpler physical systems^{51,52}, whether that boundary is a starlight-illuminated planetary surface, the bilayer-membrane which separates the inside from the outside of a living cell, or the edge of a gene-pool defined by the emergence of behaviors that treat family-offspring different from other offspring within species. Although the mix of boundary types on a given level is perhaps bewildering, the number of boundary-layers in any given hierarchy is reasonably small and well-defined.

KL-divergence also offers a mathematical template with which to inventory such layered correlations. The mathematics of pair, triplet, quadruplet etc. subsystem-correlations has for example been worked out e.g. in the context of studies on neural networks³⁸. Pair correlations go a long way to explain behaviors that look outward from a given boundary⁵³, but of course post-pair correlations may be crucial for maintaining sub-system correlations that look inward from the next boundary up.

What an understanding of the physical context tells us, however, is that we have not yet addressed the challenge of modeling a hierarchy of correlation-layers e.g. that look in & out from physical-boundaries like cell walls, tissue boundaries, metazoan skins, and molecule & idea code-pool edges. Of course the physical context described here also suggests ways to approach that analytical challenge.

For example task (or niche-network) layer-multiplicity has been suggested as a choice-multiplicity estimate of the effective number of correlation-layers in any community of social metazoans⁴⁰. The estimate is intended to model (in some monotone fashion) the physical (per capita) matchup-multiplicity maximum for that community, whose precise quantitation would require information that we don't have. Models like the one in Fig. 10, chosen to be: (i) understandable, (ii) amenable to non-invasive monitoring, and where possible (iii) correlation-nurturing, might in the days ahead allow physicists to help out in new ways.

-
- * pfraundorf@umsl.edu
- ¹ T. S. Kuhn, *The structure of scientific revolutions* (U. Chicago Press, 1970), 2nd ed.
 - ² M. Gardner, *The New Ambidextrous Universe* (W. H. Freeman, 1990).
 - ³ M. Doebeli, *Nature* **467**, 661 (2010).
 - ⁴ M. A. Nowak, C. A. Tarnita, and E. O. Wilson, *Nature* **466**, 1057 (2010).
 - ⁵ *How can 20th century simplifications benefit introductory classes?*, AAPT undergraduate education committee summer crackerbarrel (1999), cf. Tom Moore's reference to "the detail work" of content modernization, URL <http://www.umsl.edu/~fraundorf/aapt99x3.html>.
 - ⁶ Moore, *Six ideas that shaped physics* (McGraw-Hill, NY, 2003), 2nd ed.
 - ⁷ P. Fraundorf, *A one-map two-clock approach to teaching relativity in introductory physics*, arXiv:physics/9611011v3[physics.ed-ph] (1996), URL <http://arxiv.org/abs/physics/9611011>.
 - ⁸ E. Taylor and J. A. Wheeler, *Exploring black holes* (Addison Wesley Longman, 2001), 1st ed.
 - ⁹ A. Pais, *Subtle is the Lord...* (Oxford University Press, 1982), cf. note about Einstein's evolving view of Minkowski's approach on page 152.
 - ¹⁰ C. G. Adler, *Am. J. Phys.* **55**, 739 (1987).
 - ¹¹ L. B. Okun, *Am. J. Phys.* **77**, 430 (2009).
 - ¹² T. R. Sandin, *Am. J. Phys.* **59**, 1032 (1991).
 - ¹³ G. Oas, *On the use of relativistic mass in various published works*, arXiv:physics/0504111v1 [physics.ed-ph] (2009), URL <http://arxiv.org/abs/physics/0504111>.
 - ¹⁴ A. P. French, *Special relativity*, The M.I.T. Introductory Physics Series (W. W. Norton, New York, 1968), page 154: "...acceleration is a quantity of limited and questionable value in special relativity."
 - ¹⁵ T.-P. Cheng, *Relativity, gravitation and cosmology* (Ox, 2005), page 6: in special relativity... "we are still restricted to" ... "inertial frames of reference" and hence no acceleration.
 - ¹⁶ F. W. Sears and R. W. Brehme, *Introduction to the theory of relativity* (Addison-Wesley, NY, New York, 1968), section 7-3.
 - ¹⁷ W. A. Shurcliff, *Special relativity: The central ideas* (1996), 19 Appleton St, Cambridge MA 02138.
 - ¹⁸ C. Kittel and H. Kroemer, *Thermal Physics* (W. H. Freeman, New York, 1980), p. 39.
 - ¹⁹ D. V. Schroeder, *An Introduction to Thermal Physics* (Addison-Wesley, San Francisco, 2000).
 - ²⁰ C. Garrod, *Statistical Mechanics and Thermodynamics* (Oxford University Press, New York, 1995).
 - ²¹ J. Castle, W. Emmenish, R. Henkes, R. Miller, and J. Rayne, *Science by degrees: Temperature from Zero to Zero*, Westinghouse Search Book Series (Walker and Company, New York, 1965).
 - ²² T. A. Moore and D. V. Schroeder, *Am. J. Phys.* **68**, 1159 (2000).
 - ²³ M. Tribus, *Thermostatistics and thermodynamics* (D. Van Nostrand Co., Princeton, 1961).
 - ²⁴ E. T. Jaynes and G. L. Bretthorst, *Probability theory: The logic of science* (Cambridge U. Press, 2003).
 - ²⁵ E. T. Jaynes, *Phys. Rev.* **106**, 620 (1957).
 - ²⁶ E. T. Jaynes, *Phys. Rev.* **108**, 171 (1957).
 - ²⁷ P. Fraundorf, *Am. J. Phys.* **71**(11), 1142 (2003).
 - ²⁸ P. C. Gregory, *Bayesian logical data analysis for the physical sciences: A comparative approach with Mathematica support* (Cambridge U. Press, Cambridge UK, 2005).
 - ²⁹ C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, IL, 1949).
 - ³⁰ M. Tribus and E. C. McIrvine, *Scientific American* **224**(3), 179 (1971).
 - ³¹ G. D. Stormo, *Journal of Theoretical Biology* **195**, 135 (1998).
 - ³² C. H. Bennett, M. Li, and B. Ma, *Scientific American* **288**(6), 76 (2003).
 - ³³ T. M. Cover and J. A. Thomas, *Elements of information theory* (John Wiley and Sons, NY, 2006), 2nd ed.
 - ³⁴ A.-L. Barabasi, *Linked: How everything is connected to everything else and what it means for business, science and everyday life* (Plume/Penguin, 2003).
 - ³⁵ K. P. Burnham and D. R. Anderson, *Wildlife Research* **28**, 111 (2001).
 - ³⁶ K. P. Burnham and D. R. Anderson, *Model selection and multimodel inference: A practical information-theoretic approach* (Springer Science, NY, 2002), 2nd ed.
 - ³⁷ S. Lloyd, *Physical Review A* **39**(10), 5378 (1989).
 - ³⁸ E. Schneidman, S. Still, M. J. Berry II, and W. Bialek, *Phys. Rev. Lett.* **91**, 238701 (2003).
 - ³⁹ M. Harper, *Information geometry and evolutionary game theory*, arXiv:0911.1383v1 [cs.IT] (2009), URL <http://arxiv.org/abs/0911.1383>.
 - ⁴⁰ P. Fraundorf, *Complexity* **13**(6), 29 (2008).
 - ⁴¹ D. Hestenes, *Am. J. Phys.* **71**, 105 (2003).
 - ⁴² D. Hestenes, *Am. J. Phys.* **71**(7), 691 (2003).
 - ⁴³ C. J. L. Doran and A. N. Lasenby, *Geometric algebra for physicists* (Cambridge U. Press, 2003).
 - ⁴⁴ T. A. Moore, *Am. J. Phys.* **72**(4), 522 (2004).
 - ⁴⁵ E. J. Konopinski, *Am. J. Phys.* **46**(5), 499 (1978).
 - ⁴⁶ E. F. Taylor, S. Vokos, J. M. O'Meara, and N. S. Thornber, *Computers in Physics* **12**, 190 (1998).
 - ⁴⁷ C. M. Bender, *Reports on Progress in Physics* **70**, 947 (2007).
 - ⁴⁸ H. D. Young and R. A. Freedman, *Sears and Zemansky's University Physics with Modern Physics* (Addison Wesley, 2012), 13th ed.
 - ⁴⁹ S. Lloyd, in *1989 Lectures in Complex Systems*, edited by E. Jen, Santa Fe Institute (Addison-Wesley, Redwood City, CA, 1990), no. II in Lectures in the Sciences of Complexity, pp. 67–73.
 - ⁵⁰ P. Fraundorf, *Complexity* **13**, 16 (2008).
 - ⁵¹ P. W. Anderson, *Science* **177**, 393 (1972).
 - ⁵² K. Brading and E. Castellani, in *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta (Stanford University Press, Fall 2008), URL <http://plato.stanford.edu/archives/fall2008/entries/symmetry>.
 - ⁵³ W. Bialek and R. Ranganathan, *Rediscovering the power of pairwise interactions*, arXiv:0712.4397 [qbio.QM] (2007), URL <http://arxiv.org/abs/0712.4397>.